"After all, most of the 'Myth' has some blurred empirical foundation: determinants of University level performance of students: case study of Rural Development Course covered in 2008 with in Mekelle University"

Mezgebo, Taddese
Mekelle University

28. September 2008

"After all, most of the 'myth' has some blurred empirical foundation"

Determinants of university level performance of students: Case study of rural development course covered in 2008 with in Mekelle University

By Taddese Mezgebo*

Abstract

There is a widely accepted claim which insists that students' performance at Ethiopian universities is constrained by poor educational input provided at high school level and especially by students' limited capacity to communicate in English. For the specific course analyzed under this paper the impact of high school performance is found to be positive but very marginal. But based on Grade point average there is positive and significant association between high school performance and university performance, especially at freshman level. But this strong association is not specifically related to math or English performance per se but to over all performance.

*Lecturer in Mekelle University -Collage of business and economics (CBE) - Economics department – e-mail tame_new@yahoo.com and tame_new2@yahoo.com, Tel +251-911-480483, P.O.B. 451- FAX + 251-034-4407610

Determinants of university level performance of students

1. The nature and nurture question once again

The question 'is nature or nurture the determinant factor of analytical intelligence?' was and is a center of wide debate among social and natural scientists. The debate was focused on deciding if genetic codes or environmental factors are the determinant factor of human analytical intelligence. Even though it is politically correct to assume ever human being is equal to another human being in its analytical capability, it is hard to swallow such assumption in scientific arena. And at the same time it is another extreme to classify human beings in to Guardians (with higher or close to perfect intelligence) and commons (with average or below average intelligence) as is done in Plato's republic with out using an extremely rigid definition of intelligence.

The main difference is not on accepting or rejecting that people are different on their analytical capacity; since you don't need to be Albert Einstein to observe huge difference among people's capability. The main problem is in determining if genetic or environmental factors are behind such analytical difference. And if both factors are found to determine intelligence simultaneously, the next question is: which one is the dominant factor, nature or nurture?

Bouchard and McGue (1981) as sited in Nairne (1996) found that the IQ (Intelligence Quotient) of identical twins raised in different environment have high correlation of 0.72 but unrelated sibling reared together have correlation coefficient of 0.3. And half of the difference in intelligence is found to be explained by genetic difference as observed on many other studies (ibid). So it seems "similarity in environmental history is not as strong a predictor of intelligence as similarity in genetic background." (ibid: 373). Given the existence of some experimental doubts related to the reliability of the sample and given the fact that nurture and nature will normally reinforce each other which make the separation ones impact form another impossible, the conclusion must be accepted with reasonable reservation. Based on existing facts the logical conclusion is given by Nairne (1996: 376) in saying

> *The most reasonable position to take at present time is that one's intelligence, like many other psychological attributes, is determined by a mixture of genes and environment. The genes that you inherit from your parents place upper and lower bounds on intellectual ability. Genes determine how your brain is wired, and possibly the speed of neural transmission, but the expression of your genetic material is strongly influenced by environment.*

The important point for this study is answering the question 'how important is environment for average success in university?' It is well documented fact that family income, family education back ground, gender and other socioeconomic factors are important variables for success at pre-university and university level[1].

And in developing economy like Ethiopia with very under developed socio economic environment as reflected in very low adult literacy rate (35.9%), high percentage of undernourished population (46%) and high incident of poverty (23% earning below 1 USD per day) (UNDP, 2007/8) in this paper it is hypothesized that the importance of environment is much stronger than genetic on predicting university level performance. The argument goes from two sides first the level of intelligence demanded for University education is not that high to demand exceptionally intelligent students. Second high school performance measured in such unfavorable environment will not be a good indicator of intelligence and this will result on lack of link between high school exam performance and university exam performance. The justification for the second conclusion is given below.

In country where significant portion of the population living under deep socio economic poverty many flowers with good quality seed will be stunted as many poor quality flowers will blossom in good environment. Fortunately once students join university there is a large standardization of general quality of life as almost all students will use the same dormitory and dining facility. Even though difference on their socio economic status can't be denied even at university, the over all difference on their quality of life and environment will be narrowed significantly. But most importantly the university culture will encourage students to work and study hard with out being labeled geek, nerd or 'Deftera' to connote Ethiopian term. In general the over all university culture did result on higher effort exerted at university by all students and this will improve their performance. Given these changes students will be able perform to their full capacity and this will break the link between high school and university performance. At best high school performance will have just marginal impact on university performance than being the foundation of success. So it is hypothesized by the research that high school performance is just another marginal factor and there is much work that can be done behind high school at university level to improve performance of the students.

Another important issue is related to the importance of fluency in English language on good education performance at university level. Language is important parts of any inter personal communication including the teaching and learning process. But the important question is, is fluency in English language the critical factor needed for good performance at University? The point is not questioning weather effort have to be done to improve the English language capability of students or not, but the question is: is fluency in English what is critically needed for having the perfect student or it is just another factor among many important factors?. And the researcher hypothesizes that good performance in high school English is just another important but still marginal factor than the critical factor.

In this paper the two central questions state above will be asked. The first question is: is high school performance a determinant factor for university level performance? And the second question is fluency in English a basic determinant factor for good performance? Even though this paper will not provide with complete answer it is hopped that an important question and possible answer is

provided in this paper, at least from one side. For better understanding of the result of the study, let's explain the method of analysis used in this paper next.

2. Methodology

2.1. Statistical Model of analysis

In order to measure performance and understanding of students in a given course, both subjective and objective exams are often used. Objective exam will provide us with dichotomous measure of performance. And this information can be useful in identification of the determinants of students' exam performance and understanding of the course. But to use the information optimally we will need a robust statistical analysis and model which can link different independent variables with student performance.

As the use of exam is to measure students' understanding and knowledge of the course matter; it is logical to start the endeavor by asking what determines knowledge of the course. The main problem of such endeavor is related to the fact that knowledge is not observed. What we observe is exam results only. So we need to understand the link between knowledge and exam results.

The unobserved knowledge function can presented by general knowledge function of the following form, $\mathbf{k_n} = f(.)$. The vector of students' knowledge toward each exam question $(\mathbf{k_n})$ is (n) dimension vector; in which n is total number of observations (total questions × total number of students). The level of knowledge $(\mathbf{k_n})$ student has is function of student specific, input specific and question specific variables. All three groups of variables can be putted under matrix $\mathbf{X}_1$ with dimension of $(n \times k)$. In which n is as defined above and k is total number of independent variables, including intercept. Student specific variables included all variables which are specific to the student. These are his/her pre course intellectual capacity, attendance rate, level of interest to ward the course and so on. Input specific variables include the availability of text books, lecture quality, the availability of other supplementary materials and so on. And the exam specific variables are variables which determine the nature of the concept, theory or model that students are demanded to know in the exam. These are related to two facts. First, if direct or indirect questions are used and second, if subjective[2] or objective questions[1] are used in exam. These are not direct determinants of knowledge but specify what kind of knowledge is being looked at or what kind of knowledge the students are examined to have or not. Means indirectly they are determining if students are to be observed to know or not, holding other things constant. So knowledge can be presented in following general functional form

$$\mathbf{k}_n = f(\mathbf{X}_1) \quad \text{.....1}$$

Some of the determents of knowledge like the availability of text book, students' intellectual quality, attendance rate & so on can be observed. But there are also other

variables which are not observed by the researcher. These include general interest of the student to ward each topic[3], lecture quality and so on. Let's put these unobserved variables under common vector of $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_1$ is (n) dimension vector. Assuming unobserved knowledge is linear function of both observed $(\mathbf{X}_1^o)$ and unobserved $(\boldsymbol{\varepsilon}_1)$ independent variables, equation 1 can be represented in the following form

$$\mathbf{k}_n = f(\mathbf{X}_1^o , \boldsymbol{\varepsilon}_1) = \mathbf{X}_1^o\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \quad \text{.....2}$$

But what we observe is vector of exam result in dummy form $(\mathbf{D}_{er})$ not knowledge. The dichotomous exam result variables are related to knowledge by unobserved threshold level of knowledge $(\mathbf{k}_T)$. Threshold level of knowledge $(\mathbf{k}_T)$ is the minimum level of knowledge needed, if students are going to have none zero mark at given question.

$$\mathbf{k}_T = f(\mathbf{X}_2) \quad \text{.....3}$$

The threshold level of knowledge $(\mathbf{k}_T)$ is function of its determinants putted under $\mathbf{X}_2$ which is $(n \times m)$ dimension matrix. And given some elements in $\mathbf{X}_2$ are not observed the threshold level of knowledge can be decomposed in to observed $(\mathbf{X}_2^o)$ and unobserved $(\boldsymbol{\varepsilon}_2)$ components. And assuming liner threshold knowledge function it can be represented as follows

$$\mathbf{k}_T = f(\mathbf{X}_2^o , \boldsymbol{\varepsilon}_2) = \mathbf{X}_2^o\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \quad \text{.....4}$$

So exam result will be none zero, if the level of knowledge that student has is above or equal to the minimum level of knowledge that is needed to understand and answer the question right. Other wise exam result will be zero. Given $\mathbf{D}_{er}$ is a dummy of exam result, taking value of 1 if the answer is right & 0 other wise, the above statement implies

$$\mathbf{D}_{er} = 1 \text{ if } \mathbf{k}_n \geq \mathbf{k}_T$$
$$\mathbf{D}_{er} = 0 \text{ if } \mathbf{k}_n < \mathbf{k}_T \quad \text{.....5}$$

So a student will has none zero exam result in one question, if the level of knowledge that student has is higher than the threshold level of knowledge needed to understand the question and answer it right. Formally the probability if getting none zero exam result can be presented in following form.

$$P(\mathbf{D}_{er} = 1) = P(\mathbf{k}_n \geq \mathbf{k}_T)$$
$$P(\mathbf{D}_{er} = 1) = P\left[(\boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1) \leq (\mathbf{X}^0\boldsymbol{\beta})\right] \quad \text{.....6}$$

Where $\boldsymbol{\beta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ and $\mathbf{X}^0$ is a matrix of all independent variables[4] found either in $\mathbf{X}_1^o$ or $\mathbf{X}_2^o$. Equation 6 is equal to the cumulative density of the unobserved difference in determinants of threshold and actual knowledge $(\boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1)$ evaluated at $\mathbf{X}^0\boldsymbol{\beta}$. So if the

distribution of the error terms $(\boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1)$ is known, the probability of getting none zero exam result and the value of the differential coefficients $(\boldsymbol{\beta})$ can be estimated. Assuming $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ follow a joint normal distribution and they are presented under the vector of $\boldsymbol{\varepsilon}_n$, their joint density will be

$$\phi(\boldsymbol{\varepsilon}_n) = 2\pi^{-1} \left|\boldsymbol{\Omega}_i\right|^{-1/2} e^{\left[\frac{-\boldsymbol{\varepsilon}_n \boldsymbol{\Omega}_i^{-1} \boldsymbol{\varepsilon}_n}{2}\right]} \quad \text{.....7}$$

This means the vector of error terms have normal distribution with zero mean $\left(E(\boldsymbol{\varepsilon}_1) = E(\boldsymbol{\varepsilon}_2) = 0\right)$ and variance covariance matrix denoted as $\boldsymbol{\Omega}_i$. And $\left|\boldsymbol{\Omega}_i\right|$ is determinant of $\boldsymbol{\Omega}_i$. But to simplify things let's drive a combined error term $(\boldsymbol{\varepsilon})$ in which $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1$. And this will be normally distributed[5] with mean 0 and variance of σ^2 or $\boldsymbol{\varepsilon} \sim N\left[0, \sigma^2\right]$. So the density function of $\boldsymbol{\varepsilon}$ can be represented on the following form

$$\phi(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathbf{e}^{\frac{-\boldsymbol{\varepsilon}^2}{2\sigma^2}} \quad \text{.....8}$$

And replacing the new error term in to equation 6 and incorporating the assumed normal distribution, we can drive the probability function as follows

$$P\left(\boldsymbol{\varepsilon} < \mathbf{X}^0\boldsymbol{\beta}\right) = \int_{-\infty}^{\mathbf{X}^0\boldsymbol{\beta}/\sigma} \phi(\boldsymbol{\varepsilon}/\sigma)\, d(\boldsymbol{\varepsilon}/\sigma)$$

$$P\left(\boldsymbol{\varepsilon} < \mathbf{X}^0\boldsymbol{\beta}\right) = \Phi\left(\mathbf{X}^0\boldsymbol{\beta}/\sigma\right) \quad \text{.....9}$$

Equation 9 implies the fact that the probability that a student with observed above threshold level knowledge equal to $\mathbf{X}^0\boldsymbol{\beta}$ will be observed to have none zero exam result is equal to normal cumulative distribution evaluated at $\mathbf{X}^0\boldsymbol{\beta}$. And the effect of $\mathbf{X}^0$ on probability of none zero exam result can be found by differentiating the probability with respect to $\mathbf{X}^0$.

$$dP\left(\boldsymbol{\varepsilon} < \mathbf{X}^0\boldsymbol{\beta}\right)/d\mathbf{X}^0 = \left(\boldsymbol{\beta}/\sigma\right)\phi\left(\mathbf{X}^0\boldsymbol{\beta}/\sigma\right) = \boldsymbol{\beta}\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right) \quad \text{.....10}$$

The last term is obtained given the fact that for dichotomous variable the standard deviation $\left(\sigma^2\right)$ can't be estimated, so it has to be normalized to be equal to one. The next step is to estimate the coefficients of differential observed knowledge $(\boldsymbol{\beta})$. Let's take each observation as independent draw from normal distribution with probability of success equal to $\Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)$ and probability of failure equal to $1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)$. So the joint distribution of all observations can be putted in following Probit (Normit) likelihood function

$$L = \prod_{i_1=1} \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right) \prod_{i_1=0} \left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)\right) \text{.....} \quad 11$$

By maximizing equation 11, estimate of the coefficients of differential observed knowledge $(\boldsymbol{\beta})$ will be obtained. And to make things easy in actual algorithm, log of the above likelihood equation is estimated in the following form

$$LL = \sum_{i_1=1} \log \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right) + \sum_{i_1=0} \log\left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)\right) \text{.....} \quad 12$$

And the maximization of the above log likelihood function and estimation of the coefficients is done using Stata 9 soft ware.

2.2. Specification and Sensitivity analysis of the model

Two important assumptions of the above model are normality and homogenous variance of the error vector. Unless these two assumptions are assured the Probit estimates are biased and inconsistent (Verbeek 2006 and Green 2000). So the Lagrange multiplier testes for normality and heteroskedasticity are used to make sure that there are no specification biases. To test for normality lets take the following log likelihood based on more general form of distribution as adopted from Verbeek (2006)

$$LL = \begin{bmatrix} \sum_{i_1=1} \log \Phi\left(\mathbf{X}^0\boldsymbol{\beta} + \lambda_1\left(\mathbf{X}^0\boldsymbol{\beta}\right)^2 + \lambda_2\left(\mathbf{X}^0\boldsymbol{\beta}\right)^3\right) + \\ \sum_{i_1=0} \log\left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta} + \lambda_1\left(\mathbf{X}^0\boldsymbol{\beta}\right)^2 + \lambda_2\left(\mathbf{X}^0\boldsymbol{\beta}\right)^3\right)\right) \end{bmatrix} \text{.....} \quad 13$$

Taking first derivative of the above log likelihood function with respect to β, λ_1 & λ_2 will give us the necessary scores. Under null hypothesis of $\lambda_1 = \lambda_2 = 0$ the scores are

$$\frac{dLL}{d\boldsymbol{\beta}} = \sum_{i_1=1} \frac{\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}\mathbf{X}^0 + \sum_{i_1=0} \frac{-\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)\right)}\mathbf{X}^0 = \varepsilon^G\mathbf{X}^0 \text{.....} \quad 14$$

$$\frac{dLL}{d\lambda_J} = \sum_{i_1=1} \frac{\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}\left(\mathbf{X}^0\boldsymbol{\beta}\right)^{J+1} + \sum_{i_1=0} \frac{-\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)\right)}\left(\mathbf{X}^0\boldsymbol{\beta}\right)^{J+1} = \varepsilon^G\left(\mathbf{X}^0\boldsymbol{\beta}\right)^{J+1}$$

..... 15

So to test for normality none centered degree of determination of the regression of vector of ones on the above scores, times the number observations on this artificial regression will be used. This LM statistics has chi square distribution with two degree of freedom (Verbeek 2006). To test for heteroskedasticity lets develop another general form with heteroskedastic variance as adopted from Verbeek (2006)

$$LL = \sum_{i_1=1} \log \Phi\left(\frac{\mathbf{X}^0\boldsymbol{\beta}}{h(\boldsymbol{\alpha}\mathbf{Z})}\right) + \sum_{i_1=0} \log\left(1 - \Phi\left(\frac{\mathbf{X}^0\boldsymbol{\beta}}{h(\boldsymbol{\alpha}\mathbf{Z})}\right)\right) \quad \text{.....16}$$

The important assumptions in equation 16 are $h(0) = 1$, $h(\boldsymbol{\alpha}\mathbf{Z}) > 0$, $h'(\boldsymbol{\alpha}\mathbf{Z}) \neq 0$ and under the null hypothesis it is assumed that $\boldsymbol{\alpha} = 0$. Under the null hypothesis scores in equation 14 are still relevant but equation 15 will be replaced by derivative with respect $\boldsymbol{\alpha}$ and it will be in the following form

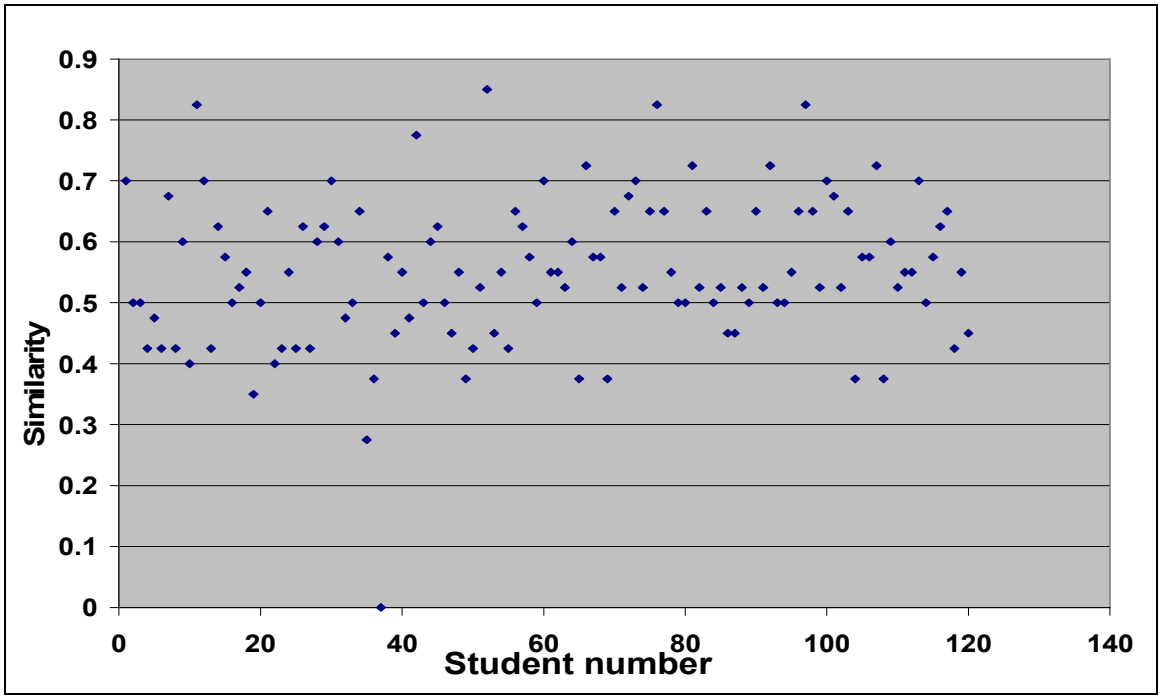
$$\frac{dLL}{d\boldsymbol{\alpha}} \left[\begin{array}{l} = \left(\sum_{i_1=1} \frac{\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}\left(\mathbf{X}^0\boldsymbol{\beta}\right) + \sum_{i_1=0} \frac{-\phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)}{\left(1 - \Phi\left(\mathbf{X}^0\boldsymbol{\beta}\right)\right)}\left(\mathbf{X}^0\boldsymbol{\beta}\right) \right)\left(-h'(0)\right)\mathbf{Z} \\ = \varepsilon_2^G \mathbf{Z}\left(-h'(0)\right) \end{array} \right] \text{...17}$$

So the none-centered R^2 of the regression of vector of ones on scores[6] given in equation 15 and 17 times the number of observations in this artificial regression has chi-square distribution with 'J' degree of freedom. 'J' is number of variables in $h(\boldsymbol{\alpha}\mathbf{Z})$ or $\mathbf{Z}$. In both cases (test for normality and heteroskedasticity), if the calculated value is higher than the tabulated value the null hypothesis is rejected and the model is assumed to be miss specified.

In addition to the above specification tests additional sensitivity analysis is needed to make sure that the exam results are actually measuring students' knowledge and understanding. The problem is if there is wide spread cheating, which can be expected in objective exams, the link between exam performance and understanding will be very weak. So the sensitivity of the analysis and conclusion for possible cheating need to be analyzed. For this purpose maximum similarity in answer between each student and any other student is calculated using a simple algorithm. The algorithm is given in appendix to the paper.

As can be seen from figure 1 below, only 4 students are having similarity of 80% and above. So even though the possibility of cheating can't be ruled out, the probability of wide spread cheating is less probable. Students having CGPA (cumulative grade point average) below 3 and which are also having higher similarities with others are dropped to see, if the conclusion is sensitive to possible cheating. The cut points of similarity are above or equal to 60%, 70% and 80% compared to any other student. But students having CGPA above 3 are not dropped in the sensitivity analysis, even if they have higher similarity with others. The assumption is that if one student with higher CGPA and other student with lower CGPA are having higher similarity the second one is cheating not vise verse. And if two high CGPA students are having higher similarity, it is assumed they are answering the question right which is resulting in similarity than cheating, per se.

Figure 1 maximum similarity between each student and any other student's exam result



2.3. Data source

In this paper rural development course offered at Mekelle University in academic year of 2007/8 for two sections is used. Most of the student and course related data are available for 3 sections. But high school level data is not available for one extension section[7]. Ignoring the high school level data can provide us with more observations. But given the wide spread claim that university level performance is related to high school level performance, specially performance in English, dropping the third section was preferred than ignoring the high school level information. In general 4218 observations from 114 students are used to analyze the issue. Moreover in order to have clear understanding of the dynamic impact of high school performance on university performance, high school level national examination results and university level grade point average (GPA) are analyzed for 373 freshman students who join to economics department in 2008/9, second year students who join to the department in 2007/8 and 106 third year students who join the department in 2006/7. Given the above facts related to methodology and data, let analyze the result, next.

3. Analysis and discussion

3.1. Definition of variables and logic for their usage

Student performance is function of inputs provided to them, their capacity and inclination to ward the course, input and effort exerted by the students themselves and other general environmental variables. Inputs provided by the university include the availability of text books, lecture quality, the availability of other supplementary materials and so on. Let's start from these variables

Table 1 variables related to input provided by the university

Variable	Description	Natural of the data
Covered in late hours	If the lecture is given in second hour	Binary
Availability of main textbook	If the part is available in the 'main' text book	Binary
Availability of other book	If the part is discussed in other general books	Binary
Availability of module	If the module of the distance program also have a related part	Binary

The course does not have a text book which can provide reference for the majority of the chapters, but they are provided with teaching material prepared by the instructor. Still different parts of the course can be found in one main text book[8], module[9] and other supplementary books. How ever given the books are either written to teach rural development from a given country prospective (India in case of main text book) or are general development text books, than rural development per se and the course is continuously reshaped to make it relevant to the countries rural development efforts, the materials are poorly adequate. As result they are provided with computer written notes in each chapter but still the material provided to them is not an ideal form. So they can still gain much by consulting the text and other supplementary books.

Other very important input is the quality of lecture provided to them. This can vary from instructor to instructor, from chapter to chapter and from day to day. But data on instructor quality is not available and it is treated as unobserved determinant of knowledge[10]. Fortunately a related data is available on timing of lecture. From the three credit hours per week, two of them are covered by two hour class. As result 1/3 of the over all lecture is given in the second hour. And it is an important research question to identify, if students performance is related to duration of the lecture hours. Students and the lecturer will be exhausted some how after one hour class and it is expected that both lecture quality and students understanding and attention will go down. And this will be also reflected on exam performance.

The second important group of variables is related to the intellectual capacity and inclination of the student to ward the course after joining university. The capacity of the student can be measured by the pre course cumulative grade point average (CGPA). Moreover by using attained grade on related course not only the capacity of the student but also inclination to ward similar courses can be measured (though imperfectly). Given

the fact that rural development has higher overlap with development economics and agricultural economics their performance on these courses are controlled. But the coding of grades is made by giving 1 to A, 2 to B and so on, so higher value show poor performance.

Table 2 variables related to capacity and inclination of the students

Variable	Description	Natural of the data
Grade on development 1	Grade on development economics 1	Ordered from best to worst in ascending order 1 – A, 2 –B, 3 –C & 4-D
Grade on development 2	Grade on development economics 2	
Grade on agriculture economics	Grade on agricultural economics	
CGPA	Cumulative grade point average	Scale in range of 0 to 4

And it is well documented fact that well motivated and competent students will perform much better at high school level (Amare 2001). Study by Stephenson et al. (2006) among others did also find that students' CGPA is a good predictor students' performance in university level courses, too. So both logic and empirical evidence did support the incorporation of the above variables.

The widely believed view about students' performance at Ethiopian university is that students perform better, if they have good performance at high school and especially if they are fluent in English language. Such causation is also well documented in many countries[11]. In most case studies good performance not only on high school courses in general but also specifically on math and English are found to be an important predictors of university level success.

Table 3 Measuring high school performance of the students

Variable	Description	Natural of the data
English result in ten	English grade in Ten	Ordered from best to worst in ascending order 1 – A, 2 –B, 3 –C & 4-D
Math result in ten	Math grade in Ten	
Share of grade A in 10	Share of A grade in ten grade exam	Scale from 0 to 1
Share of grade B in 10	Share of B grade in ten grade exam	
English ESLCE result	English result in grade 12	Percentage
Math ESLCE result	Math result in grade 12	Percentage
Average ESLCE result	Average percentage at 12	Percentage

However even though students which perform better in freshman will have much higher probability that they will also perform better in senior years; the comparative disadvantage of those who perform poorly in freshman will be eroded quickly with years[12]. The implication is that students which are not performing well at high school are not necessary low IQ students. It is possible that higher IQ students can perform poorly due to negative environmental and social factors and influences[13]. So the change in

educational, meal and housing environment which is introduced with university level education can have a strong positive impact on bringing what is hidden with in. To account for such effects high school level performance on both math and English is measured by English and math result in both grade 10 and grade 12. Moreover the over all performance in high school is measured by share of A's and B's in grade 10 national examinations and average ESLCE result taken at grade 12. Both grade 10 and grade 12 exams are given at national level, so they will provide us with robust measure of high school performance.

Other important group of variables is related to level of input exerted by the students' themselves. These can be represented by attendance of the lectures provided and study hours allocated to the specific course. At high school level study done by Mathewos (2000) did find that students which allocate longer time for study and which are not missing frequently due to health related reasons[14] did perform better than others.

Table 4 Measures of students' attendance rate

Variable	Description	Natural of the data
Specific attendance	If the student was in attendance when the lecture related to the question is covered	Binary
Explained in other parts	If the concept related to the question is also lectured in other day	Binary
Attendance on other related class	If the student was in attendance when the concept is covered in second date	Binary
Attendance rate of the chapter	Attendance rate of the chapter related to the question	Scale from 0 to 1
Total attendance	Attendance rate of the course	Scale from 0 to 1
Change in attendance	Change in rate of attendance after mid	Scale from 0 to 1

For this paper data is collected on daily attendance of each student and course parts covered at each lecture day. Given these data, attendance rate of each student when each theory or concept related to each question is lectured is derived to measure the impact of attendance on performance. Moreover given the fact that concepts in any chapter and the entire course are interrelated, students' level of understanding is also expected to be effected by the attendance rate of the chapter and the entire course. Additionally some theories are dealt many times given interrelated nature of the chapters[15]. So control variables are introduced to see if the concept is covered in other parts and if the student was in attendance at such date. Unfortunately data is not available in students study hour allocated to the course, so it is taken as unobserved determinant.

But one additional but very important variable is change in attendance rate observed after mid. Two things happen after mid term exam. First most of the students perform poorly in mid term exam and students were stressed to complete their senior essay as graduation is less than month and 30 days away. So observing poor performance in the mid term exam result, students made a decision to cut attendance of the course given the intense shortage of time they are facing. The change in attendance rate or difference between pre

and post mid attendance rate will be help full in explaining students' performance by level of stress and rationality. Insignificant and even negative effect will show high level of rationality. Means having shortage of time and their information about importance of lecture on exam performance their decision to cut attendance rate is optimal. This is so because significant negative coefficient implies that students which reduce their attendance rate have better probability of performing well in exam. Means they were more rational than others. And if the effect is insignificant what ever decision that is done by all students was rational.

But significant positive coefficient does not necessarily show irrationality, as it can be related to stress. From one point of view, it can mean that students by cutting attendance rate they were able to perform poorly, which reflect irrationality. But from other point of view given that they were facing shortage of time they try to minimize lose by making sure their senior paper is ready, even if it means having low grade in other courses. Means students which are performing well when their attendance rate is increasing are also students which are not stressed due to shortage of time. So they may be doing better not because they are attending the lecture but because they are less stressed.

Even though the above variables are the most important variables (for this paper), it is still important to know if performance is related to gender and region of origin. Gender is obviously important and the impact of gender on performance is widely documented. To give some examples in Ethiopia high school level data analyzed by Mathewos (2000) and Amare (2001) found that boys have better performance than girls. In Bangladesh study by Asadullah et al. (2007) found that girl high school students were under performing compared to the boys. And more robust study based on international data from 266,545 secondary-school students from 39 countries by Wößmann (2003) found similar result. So most studies are showing the fact that at high school level boys are normally out performing girls. At elementary level study by Ahmed et al. (2006) in Bangladesh found that girls were under performing in math but not in Bengal (local language). So both elementary and high school performances of girls were poorer than boys. From another side Bacolod, et al. (2006) based on data from the Cebu Longitudinal Health and Nutrition Survey (CLHNS) Philippines, show that girls were more effective in improving their performance than boys; but he did not reveres the widely held view that boys will perform better than girls. At university level Stephenson et al. (2006) found that being male have positive impact on analytical performance. And the impact of gender will increase with age. The same result is found by Nolan et al. (2007). How ever Salahd, et al. (2005) did not found any significant impact of gender on first year students' performance. So most of the evidence is directing to ward the fact that in all levels girls will perform poorly than boys and the reason given include the following among other things

1. A study by Herbert et al. (2005) found that parents action and behavior which explicitly accept boys as better off will result on low self image of females as student and inferior carrier choice as adult. Parents were observed to exaggerate the achievement of boys but down grade the achievement of girls. This will result on low self image of girls as child. To make things worst the impact of family behavior on

girls' low self image is observed to increase with age. If parents were effective in creating low self image this can result on low motivation and poor performance by girls.

2. Moreover given low expected benefit from women, who normally leave their parents house after marriage, families are highly motivated to invest in their son's education than daughter's education. Moreover imperfection in labor market which disfavor women will create negative incentive to ward investment in girls' education. And this is proofed indirectly by Beutel et al. (2002) study of Nepal which shows the fact that being girl increase the probability of dropping out and will reduce the probability of enrolling to school
3. "Gender differences in educational attainment may arise because of gender differentiation in adult roles and the emphasis on family-related roles for women" (Beutel et al. 2002: 110). This is shown by their study result in which leaving school for search of job is higher among boys than girls. But dropping out due to marriage is more prominent among girls than boys.

The impact of negative self image and condition which is created at family and local level though expected to diminish after joining university; it will not be expected to evaporate just like that. And this is proven by study of Fentaw (2001). Fentaw's research result showed the fact that among 289 female students which join AAU faculty of social science in 1994, 54% of them did not complete their course in 4 years uninterrupted and after 6 years the over all graduation rate is around 60%. This is low achievement by any measure. And this is due to two reasons first as the problem is structural any changes in environment will not eliminate the entire problem in short period of time. Second university will come not only with its own solutions but also with its own problems.

A study by Asresash et al. (2002) in Jimma University, for example, found that 10% and 27% of the female students did admit to be harassed by staff and students, respectively. And the sampled boys and girl students did agree on the fact that girls' performance was also constrained by poor educational back ground, lack of confidence, lack of capability to manage the new obtained freedom and others. So their early poor back ground and other challenges added at university level will result on poor female students' performance, especially at freshman level. But behind freshman the performance and survival of female students is observed to improve (Fentaw 2001). And this is not special to female but for all groups which are under prepared to university level education, as was stated before.

Table 5 General and control variables

Variable	Description	Natural of the data
Sex	If male the value is 1 and 0 other wise	Binary
Region	Region in which high school is completed	Multinomial
Class level mark of the Question	Average mark of all students in each question	Scale from 0 to 1
Chapter	Dummy for each chapter	Multinomial

Regional location where high school is completed is an important variable for controlling the impact of region level factors like: regional education policy, socio-cultural and socio economic environment. Studies did widely document the fact that students from better off family or better off areas or better off countries have better performance at pre-university and university level[16]. Moreover students which are leaving in family having members with higher educational attainment did perform better than others[17]. So the impact of local socio economic and socio cultural environment will be controlled to limited extent by regional location.

Additionally school quality how ever it is measured is found to be an important factor in explaining students' performance. Study by Nolan et al. (2007) did found that students coming to University of Sydney form private school did perform poorly compared to those who come from public school. And study by Asadullah et al. (2007) in rural Bangladesh high schools did show that students which complete their primary education in religious school did perform poorly in high school compared to others. Moreover schools which are awarded for their good quality did generate students with better exam performance compared to others. Additional evidence from Bangladesh by Ahmed et al. (2006) and from Philippines by Bacolod et al. (2006) found that good quality school and specially those schools which have electricity resulted on good students performance compared to others. So even though there are disagreements and doubts in how to measure teachers' quality; if expenditure per school and per student are important measures of school quality; if class size and student per teacher are reflection poor quality or not; there is one widely accepted fact[18]. That is school quality in its totality matters for good performance of the students. So even though region is poor proxy for difference on school quality; given the establishment and administration of elementary and high school is under regional state autonomy, region is used as control group in the study.

Moreover given three regions (or city administrative councils) are highly urban we can use them to see if being urban have impact on students performance as documented in Sander (1999), Wößmann (2003) and others. But region will not be expected to be a good proxy for controlling difference in socio-economic, socio-cultural and educational service supply in sub regional dimension. This include difference in local education quality, local culture, family level income, family educational attainment and soon. These differences will be considered as unobserved determinants of performance, given lack of data.

Additionally, it is observed in some studies that younger students are more probable to perform better than older students. To give example Mathewos (2000) analysis on high school around Addis Ababa and Salahdeen et al. (2005) analysis on students admitted into the University College of Medicine, Lagos in 1998/99 found that older students perform poorly compared to younger students. This could imply that students will have problem of concentrating in school with increase in age and associated increase in adult responsibilities. But for this study fully compiled age data was not available so age is also treated as unobserved determinant of performance.
Additionally control over unobserved determinants which are common to all students is done by including average performance of all students in each question. So the real dependent variable is not students' performance per se, but deviation from average

performance. And given the fact that important variables like lecture quality are not controlled this will be an important variable in avoiding specification problem related to exclusion of important variables. One more additional group of variables which are found to be important for achievement of normality is dummies for each chapter. Given the above facts now we are ready to analyze the result.

3.2. Analysis of model result

Maximum likelihood statistics are asymptotically efficient and consistent but have unknown small sample properties, in most cases (Verbeek 2006 and Green 2000). For this model 4218 observations are used which are adequate for accepting the maximum likelihood results based on asymptotic theory. The variables which are fitted did result on log likelihood of - 2026.55, which is 24% larger than log likelihood with intercept only. And all independent variables are having significant impact on exam performance at less than 1% level. Moreover the null hypothesis of normality is not rejected up to 22% level of significance. Means the assumption of normally distributed error term which is a corner stone for the model specification is acceptable assumption.

Table 6 Sensitive analysis for the main model

Variable	Value	Variable	Value
Number of observations	4218	LM (normality)	3.0181667
Log likelihood	-2026.55	chi2(2)	0.22111257
Wald $chi^2(37)$	1091.10		
Probability > chi2 (37)	0	LM (heteroskedasticity)	42.966582
Pseudo R^2	0.2420	chi2(38)	0.26684216

To test heteroskedasticity the relationship between knowledge and threshold level of knowledge is used. After reviewing result of different studies, psychologists did conclude that intelligence quotient (IQ) which is directly related to analytical inelegance needed in schools has normal distribution (see Anastai 1961 and Nairne 1996). So there is logical foundation to assume the unobserved knowledge is also normally distributed. If students' knowledge is normally distributed with constant variance, heteroskedasticity have to be introduced from threshold level of knowledge. In areas with adequate inputs students may be asked tough questions to get 1 mark but in area with input problem they may be given slack. So the threshold level of knowledge will have different distribution depending on inputs. As result observed inputs given by the university are included as the determinant of standard deviation. But to control for what ever left over effects chapter and question dummies are also included. Moreover to control for possible specification of exams in relation to class attendance, class level attendance of all students in each theory and chapter is used. In general 49 variables are tried to explain heteroskedasticity but 12 are dropped due to multicollinearity so the artificial heteroskedastic regression is fitted with 37 variables.

And as can be seen above the null hypothesis of constant variance is not rejected up to 26% level of significance. Given the above facts it is safe to conclude that the model is appropriately specified except for possible problem of omitted variables. One possible omitted variable is student dummy which can control for student level fixed effect. But

including such student dummy resulted on lack of normality on error terms and also introduced heteroskedasticity problem. Given Probit model is useless unless the assumption of normality and constant variance is satisfied (Verbeek 2006 and Green 2000), the model with out student dummy is accepted as best possible fit. Now let's analyze the marginal effect of the independent variables on exam result one after the other.

Table 7 The impact of observed university level inputs

Variable	Marginal effect	Standard error	T value	probability	Mean value
Covered in late hours	0.008	0.018	0.47	0.637	0.304
Availability of main textbook	0.022	0.042	0.51	0.607	0.342
Availability of other book	0.031	0.028	1.11	0.266	0.605
Availability of module	-0.027	0.041	-0.66	0.51	0.5

Table 7 clearly shows that there is no significant difference on exam performance, if the chapter is covered in the 'main' text book or other general development books or neither of them. Taking the numeric coefficients at face value, excluding distance model the impact of other materials is found to be positive. But the impact of each of them is not statistically significant even at 26% level and this is inline with the earlier assumptions that the text books are not adequate to provide deeper support for students. The assumed text book is highly Indian centered and does not focus on general theory. In fact the theories are simply approached from Indian prospective only. Moreover the distance module which more often mirrors the 'text' book is not found to have any significant impact. Theoretically other related books can be helpful in improving students understanding; but when students are facing series shortage of time consulting such general development books can be counter productive than productive. Moreover in contrast to the normal expectation it does not seem to matter for performance, if the concept is discussed in first or second hour of the lecture. Means having two hours class than 1 hour class did not have any significant impact on exam performance. Having longer classes even if it has no impact on performance, it may have negative impact on students comfort.

Note that all the above variables are binary and such marginal effect analysis is not appropriate for binary variables. But in many applications the result which can be observed from the right kind of analysis[19] is found to be much closer to such marginal analysis (Green 2000). Given the above well accepted wisdom, in this paper binary variables are treated as any continuous variable.

Table 8 the impact of student intellectual quality and inclination

Variable	Marginal effect	Standard error	T value	probability	Mean value
Grade on development 1	-0.029	0.018	-1.59	0.112	2.676
Grade on development 2	0.006	0.019	0.34	0.737	2.342
Grade on agriculture economics	-0.035	0.013	-2.77	0.006	2.468
CGPA	0.055	0.039	1.39	0.165	2.618

CGPA will measure the students' intellectual quality as measured by university level exam. But more importantly students' interest and motive to ward the course in combination with their intellectual quality can be imperfectly measured by their achievement in other related courses. The related courses are development economics and agricultural economics. When students' cumulative grade point average (CGPA) increase by 1 unit, the probability of answering the question right will increase by close to 5.5%. But it is not significant at conventional 1%, 5% or 10% level, but only at more than 16.5% level. Students who perform better in agricultural economics have higher probability of performing well in rural development, too. One grade point improvement in agricultural economics grade will result on 3.5% more probability of answering the question right. But the impact of development economics is mixed. Good performance in development economics one has positive impact on performance in which one level improvement in development economics one will result on close to 3% improvement in probability of getting the answer right. But it is only significant at more than 11.2% only and the impact of development economics two is not only having wrong sign but also is highly insignificant. So even though rural development seems to be more related to agriculture economics than to development economics, it is clear that performance on related courses are interdependent showing the importance of students' inclination and interest for good performance.

In general students' inclination, interest and intellectual quality are important determinants of their performance. But the million dollar question is how dependent is university performance on high school performance. As was stated above performance at high school ten grade national examination and 12th grade ESLCE national examination results are used to analyze the impact of these variables.

Table 9 High school level performance and performance in rural development

Variable	Marginal effect	Standard error	T value	probability	Mean value
English ESLCE result	-0.002	0.002	-1.23	0.218	51.108
English result in ten grade	-0.02	0.02	-1.03	0.304	2.234
Math ESLCE result	-0.002	0.001	-1.84	0.066	44.279
Math result in ten grade	0.066	0.015	4.49	0	3.099
Average ESLCE result	0.009	0.003	2.58	0.01	54.909
Share of grade A in 10	-0.049	0.086	-0.57	0.572	0.111
Share of grade B in 10	0.037	0.064	0.59	0.558	0.289

Students who perform better in high school as measured by their average mark at Ethiopian School Leaving Certificate Examination (ESLCE) have higher probability that they will also perform better at rural development course and the impact is significant at 5% level. But the marginal impact is very low in which one percentage point increase on ESLCE average result will result on 0.9% percentage increase in probability of answering the exam question right. Means 10 percentage points increase in average ESLCE result will increase the probability of answering the question right by 9%. And notice that increase by 10 percentage points on average ESLCE result can make a huge difference in the ability of student to join or not to join university. Given this result: is it rational to accept the claim that says 'university performance is all about high school performance'?

Good performance at Math of grade 10 and 12 is observed to have negative impact on good exam performance on rural development, against the widely accepted view to contrary. If we compare a student with A in ten grade math against other students with B in ten grade math, holding other things constant, the second person have 6.6% more probability he/she will perform better in rural development. And the impact is significant at 1% level. Taking 12^{th} grade ESLCE math result, one percentage point increase in ESLCE math result will reduce the probability of answering rural development question by 0.2% and the impact is significant at 6.6% level.

If we accept the theory that students inclined to ward arithmetic are less efficient with courses which need memory and vise verse, it could make sense. But from other angle students with better capacity in math are expected to be better in articulation. So the problem can be lack of time which resulted on memorized reading than articulation, in which students inclined to ward arithmetic than memorization will be at disadvantage.

The impact of high school English on performance is not only mixed but also highly insignificant. In general high school performance on general, and math and English performance on particular seems not to be the most critical factor for good performance on rural development. And observing the much emphasis given to high school performance as main determinant of university performance on formal and informal discussions of the university community and the contradictory result given above the researcher was inclined to conclude that most of the popular believe is based on unproven 'myth' [20].

But like any statistical or econometric model the above result is based on many assumptions which are used in developing the model. The general assumption is that there is no misspecification due to heteroskedasticity of variance, non normality of the error term or omission of important variables. Even though the first two are formally proved the importance of omitted variable can't be ignored so a simple and partial correlation coefficients between high school and university performance are given in table 10 below.

Table 10 correlation analysis with total exam mark

	Simple correlation				Partial correlation		
Variable	r	Significance	r^2		r	Significance	r^2
CGPA	0.534	0	0.285				
Average ESLCE result	0.35	0	0.123		0.335	0	0.112
Math ESLCE result	0.072	0	0.005		-0.212	0	0.045
English ESLCE result	0.233	0	0.054		-0.08	0	0.006
Share of grade A in 10	0.142	0	0.02				
Share of grade B in 10	0.031	0.045	0.001				
Sum							0.163

Observing the simple correlation coefficient, the most important variable for performance seems to be cumulative GPA in which with out controlling for other variables it explains

more than 28% of the variability on rural development performance. The partial correlation between performance and CGPA (which is not shown above) after controlling high school 12th grade total, math and English performance is 0.4438 means it explains close to 20% of the variability on performance. This is followed by average ESLCE performance which explains nearly 12% of the variability on performance. And with out controlling for impact of other variables good performance in high school English explained more than 5% of the variability in the course performance. And the related impact of high school math is less than 0.6%. At the same time good performance on ten grade national examination though have numerically small impact the impact is positive and significant up to 5% level. So the above simple result seems little bit at add with the Probit model result. But observing the partial correlations we can realize the fact that the most important factors for good performance in the course are high CGPA and better high school performance than good performance on math and English, per se. This is so because the impact of math and English is not only weak but also is having wrong sign when overall high school performance is kept on average on the partial correlation analysis. So it seems if high school performance has any positive impact on university performance, its impact is not highly strong; specially compared to the widely held view which takes it as central determinant of performance since it explain only less than 13% of the variability of rural development performance, at best.

But the result found for one course must be taken with reasonable reservation for three reasons. First good performance on English language and better capacity to do mathematical articulation may be more important for subjective exam than objective exams used in the course. Second the sample may not be representative of the whole population. But most importantly doubting the above result is acceptable given the huge international empirical evidence to the contrary. To eliminate the impact of un-representatives of the sample and the course simple and partial correlation analysis is done with CGPA of the students. And with out controlling for other variables ESLCE result is observed to explain around 26% of the performance on university with simple correlation coefficient of more than 0.51. And good performance in English is related to nearly 15% of variability in CGPA with out controlling for other variables. Math performance at 12th grade is positively associated with nearly 7% of the variability on university performance. Similarly 10th grade performance is associated with close to 10% variability in university performance. So even though high school performance is not the only important factor, still it and all related factors[21] are important factors for good performance at university level.

Table 11 correlation analysis with CGPA

	Simple correlation			Partial correlation		
variable	r	Significance	r^2	r	Significance	r^2
Average ESLCE result	0.516	0	0.266	0.318	0	0.101
Math ESLCE result	0.258	0	0.067	-0.059	0	0.004
English ESLCE result	0.386	0	0.149	0.054	0	0.003
Share of grade A in 10	0.279	0	0.078			
Share of grade B in 10	0.159	0	0.025			
Sum						0.107

But the much emphasis given to math and English seems unwarranted. This can be observed from the partial correlations in which keeping math and English performance on average high school performance explained more than 10% of the variability on university performance. But controlling for over all performance the impact of good performance in high school English though positive is less than 0.2% and the impact of 12^{th} math is negative and negligible.

This is why the researcher was inclined to conclude that the much emphasis given on high school performance as main and the only determinant factor for good performance at university is unnecessary. Especially the much emphasis given for fluency in English and superior understanding of math is based unproved 'myth'. But the question is why is 'myth' highly advocated by almost all stake holders? Is there unobserved factor in the above analysis? And the answer is yes. This is related to dynamic impact of high school performance from freshman until graduation. Studies stated in last part did show that high school performance is very important for freshman students' success but its importance and the comparative disadvantage of unprepared freshman students will decline with time. Such dynamic analysis can possibly show us the missing clue for understanding the foundation of the 'myth'.

Observing table 12 below, we can see how important high school performance is for freshman success. The GPA of freshman first semester 373 economics students who join at 2007/8 to economics department has 0.73 correlations with high school ESLCE total result. Means variables related directly and indirectly to high school performance explain more than 53% of the variability in freshman performance. And if we use data from 102 students who join the department at 2007/8 the importance of high school and related variables for freshman performance is more than 41%.

Table 12 Correlation of high school performance and university GPA at freshman

For first year first semester GPA of 373 students who join at 2008/9						
Variable	Simple correlation			Partial correlation		
	r	Significance	r^2	r	Significance	r^2
Average ESLCE result	0.733	0	0.538	0.484	0	0.234
English ESLCE result	0.608	0	0.37	0.111	0.033	0.012
Math ESLCE result	0.322	0	0.104	-0.064	0.221	0.004
For first year first semester GPA of 102 students who join at 2007/8						
Average ESLCE result	0.64	0	0.41	0.53	0	0.281
English ESLCE result	0.301	0	0.09	-0.224	0.025	0.05
Math ESLCE result	0.408	0	0.167	-0.115	0.253	0.013

What is surprising is that in all cases the partial correlation of math ESLCE result is not only negative but also insignificant. But the impact of English is mixed but English also seems to be less important. What is needed is not good performance at high school math and English, per se, but better over all performance at high school level at least for success in freshman of economics department. This is so given even after keeping high school math and English performance at average; good high school performance and

related variables explain more than 23% of the success in freshman performance but not vise verse. Additional and interesting clue about the dynamic impact of high school performance on university performance is observed, if second semester data is analyzed.

If we observe the correlation between high school performance and second semester performance of the 102 students stated in table 12 above, as given in table 13 below. The uncontrolled importance of over all performance at high school becomes less than 30% from more than 41% observed just one semester earlier. Means high school performance and all associated variables are explaining less than 30% of the second semester success of freshman students'. Observing math and English with out controlling other variables they have strong positive association with performance. But the question is: does good performance on math and English or general over all performance is what matters the most? To see that partial correlation are fitted and as can be seen from partial correlation given in table 13 below the important factor is better performance on high school in general not better performance on math and English, per se.

Table 13 Correlation of high school performance and GPA behind freshman

For first year second semester GPA of 102 students who join at 2007/8						
Variable	Simple correlation			Partial correlation		
	r	Significance	r^2	r	Significance	r^2
Average ESLCE result	0.545	0	0.297	0.298	0.003	0.089
English ESLCE result	0.389	0	0.151	0.055	0.586	0.003
Math ESLCE result	0.347	0	0.121	0.01	0.922	0
For first year second semester CGPA of 107 students who join at 2006/7						
Average ESLCE result	0.394	0	0.155	0.109	0.268	0.012
English ESLCE result	0.245	0.011	0.06	0.069	0.488	0.005
Math ESLCE result	0.393	0	0.154	0.219	0.025	0.048

And if CGPA of first year second semester 107 students who join at 2006/7 is considered the fact that the importance of high school performance and all related variables is less than 16%. But the important subject seems to be math among all high school courses for the batch of 2006/7. This is contradictory to most of the results and it is clear indicator of the importance of having a representative sample. Still even though the above correlation does not mean causation, what is sure is that the importance of high school performance will decline over time as students survive the first semester. This can be related to two facts. One possible explanation is that the less capable students will drop out on freshman to reduce the impact of high school performance. Other explanation is the de-conditioning and improvement in capability that can be injected at university level will result on better performance by less capable students. And such effects are also observed in other case studies. So if the second hypothesis is right it may take time before students learn to use their full capacity and before full reconditioning is done by university culture and environment to build their confidence and increase their effort. An important point is that observing the above contradictory conclusion reached about importance of high school level competence based on one course performance and CGPA performance; it does not need further proof to stress the importance of representative sample.

Table 14 Attendance and performance

Variable	Marginal effect	Standard error	T value	probability	Mean value
Specific attendance	-0.02	0.029	-0.69	0.492	0.563
Explained in other parts	-0.056	0.043	-1.31	0.191	0.6
Attendance on other related class	0	0.028	0.01	0.993	0.158
Attendance of the chapter	0.017	0.048	0.35	0.724	0.028
Total attendance	0.031	0.06	0.51	0.608	0.566
Change in attendance	0.103	0.043	2.39	0.017	-0.529

The next issue is related to the question, if attendance of lecture matter for students' performance or not? As can be seen from table 14 above, at least for the specific course attendance seems to be unproductive, if not counter productive. If students attend when the concept or theory related to the specific question is lectured, they seem to have more than 2% probability that they will do poorly in exam.

This is puzzling because logically at least you will expect attendance to be unimportant at worst, and to have positive impact at best. Fortunately, the negative coefficient is insignificant up to 49% level, so it can't be sure if the result is due to random error or not. And controlling for repetition of lecture did not produce significant impact. The logical conclusion that can be drawn from the fact is that attendance rate does not seem to be an important variable for student performance. Total attendance rate of the course and the chapter have positive impact. But still all are not significant even at 60% level. And this will reinforce the earlier conclusion that there is no statistical evidence that show attending a specific theory or over all attendance of the chapter or the course to have any impact on exam performance. This is in line with rational student (agent) theory which assumes that students which are attending are attending because they are better off by doing so and those who are not attending are also doing so because they can do better with out attending the course. If students have full information and are rational enough they will weight the time they have to lose on attending the lecture and the level of information they can lose by missing the lecture. And given high difference between students, rational students will chose what is best for them and as result attendance, holding other things constant, will have insignificant impact on performance.

However students which increase their attendance by one unit after mid will have more than 10% probability that they will answer the exam question right (table 14, above). And this effect is significant at 5% level. This could imply irrationality in part of the students. Means their decision to cut class after mid was irrational. But observing the previous fact, in which attendance is observed to have insignificant impact, it is more probable that students which are increasing their attendance or are reducing it by small margin are those with less time shortage. Those who increase their rate of attendance and which are performing better at the same time, are doing well not because they are attending but because they have enough time to read, to attend class and stay relaxed. How ever still the two main conclusions drown here have to be proved by different studies which deal with different exam modes, lecturer, courses and students before getting acceptance as a fact.

From this study point of view, how ever, observing positive impact of increase in attendance when attendance rate itself has insignificant impact and observing students which are mathematically inclined are also performing poorly, it is rational to conclude shortage of time and the associated high level of stress are exerting strong negative impact on performance. The implication is that the push for strict attendance of students on all days may be counter productive. This is against the widely accepted view and finding by Amare[22] (2001) which conclude that strict teacher control on student behavior and attendance is important for good performance.

The last groups of variables which need to be discussed are presented in table 15, below. An important variable is the class level performance. It is found that if the class is performing well the student will perform better as well. And the impact is statistically significant at 1% level. So the impact of unobserved common variables like lecture quality and other variables which effect performance are controlled by introducing average performance of all students in each question. And given the numerically and statistically significant coefficient observed it is logical to assume most common factors are controlled by the above common control variable.

Table 15 the impact other independent variables

Variable	Marginal effect	Standard error	T value	Probability	Mean value
class level mark of the question	1.121	0.04	28.06	0	0.663
sex	0.024	0.038	0.63	0.529	0.928
Afar	-0.176	0.097	-1.81	0.07	0.009
Amhara	0.033	0.027	1.23	0.219	0.117
Oromia	-0.035	0.021	-1.63	0.102	0.234
Somalia	-0.017	0.045	-0.39	0.7	0.045
Benishangul-Gumuz	0.139	0.065	2.15	0.032	0.009
SNNPR	-0.027	0.036	-0.77	0.439	0.063
Harari	-0.227	0.089	-2.56	0.011	0.009
Addis Ababa	-0.049	0.104	-0.47	0.638	0.009
Dire Dawa	-0.023	0.046	-0.49	0.626	0.036

Observing gender effect, boys are more than 2% more probable to answer the question right than girls, but the difference is statistically insignificant, even at 52%. So the impact of gender difference on exam performance of the course is statistically negligible, which goes against the orthodox wisdom. So the paper will be along the few exceptional papers which find insignificant gender impact. How ever given the students are senior and the early evidence which shows that the negative impact of gender will decline with time, it is also consistent with wide empirical evidence.

To analyze the impact of regional location students which graduate from Tigray high schools are taken as control group. Students which graduated from Benishangul-Gumuz high school have more than 13% higher probability to perform better in exam and the difference is significant at 5% level. This is another surprise given the fact that to join university students from Benishangul-Gumuz, as other under developed region students,

are given preferential treatment under affirmative action. So you will expect them to perform poorly. But another region from which students will able to join under affirmative action, i.e. Afar is having negative impact which is significant at 10% level. Students from Oromia and Harari did perform poorly in the course. And respectively they are more than 3.5 % and 23% less probable to answer the question right. The impact of regional location being in Harier is significant at 5% level but the impact of Oromia is only significant at more than 10% level.

Any clear picture does not seem to emerge from the regional location analysis. Students from one underdeveloped region are doing well, but those from other underdeveloped region are doing worst; the two regions with higher urban population are not different from majority but Harari is among the three worst. And Oromia is too big to provide any logical reason except for possibility of poor education policy and implementation in the region. However given students from Oromia are 24% of the students and given the regional impact is insignificant at conventional 1, 5 and 10% level, it can cast doubt if the Oromia's coefficient is significant at all. One additional group of variables included is a group of 6 chapter dummies. However their impact is observed to be insignificant as the Walda statistics is 4.74 and this statistics is significant at 57.82% level only. They can be dropped with out significant lose of information, how ever when ever dropped both none normality and heteroskedasticity problems are observed, so they are kept in, though they explain nothing.

3.3. Sensitivity analysis for distortion from any possible cheating

The over all model specification was done under the assumption that exam marks are measuring above acceptable threshold level of understanding. But if there were wide spread cheating the link between knowledge and exam mark will be completely distorted.

Table 16 Specification test for all models

	Level of similarity accepted for those below CGPA of 3			
statistics	100%	< 80%	< 70%	< 60%
General fitness test				
Observations	4218	4142	4066	3116
Log likelihood	-2026.5491	-1995.16	-1958.61	-1503.25
Wald chi2(37)	1091.1	1074.26	1055.66	827.1
Prob > chi2	0	0	0	0
R^2	0.242	0.2429	0.2442	0.2524
Normality test				
LM statistics	3.018167	2.726301	1.769732	0.53264
chi2(2)	0.221113	0.255853	0.412769	0.766194
Heteroskedasticity test				
LM statistics	42.966582	42.52039	40.19141	47.78663
chi2(37)	0.26684216	0.282605	0.373384	0.13272783

So sensitivity analysis is done by dropping students which are having higher similarity with others despite low cumulative grade point average. And under all considered restrictions, the models are having independent, identically and normally distributed error

terms. And the null that all coefficients are jointly zero is rejected in all cases at 1% level. So unless for any possible omitted variable bias, it is rational to conclude that all the models are well specified to link the independent variables with exam performance.

Table 17 Sensitivity analysis for possible cheating

	Level of similarity accepted for those having CGPA below 3							
	100%		< 80%		< 70%		< 60%	
Variable	Coeff.	Prob.	Coeff.	Prob.	Coeff.	Prob.	Coeff.	Prob.
Covered in late hours	0.008	0.637	0.011	0.53	0.011	0.536	0.027	0.196
Availability of main textbook	0.022	0.607	0.029	0.487	0.034	0.431	0.013	0.798
Availability of other book	0.031	0.266	0.031	0.272	0.029	0.317	0.037	0.271
Availability of module	-0.027	0.51	-0.03	0.47	-0.026	0.537	-0.008	0.866
Specific attendance	-0.02	0.492	-0.025	0.403	-0.015	0.622	-0.015	0.665
Attendance of the chapter	-0.056	0.191	-0.047	0.285	-0.068	0.128	-0.049	0.352
Explained in other parts	0	0.993	-0.002	0.937	0.001	0.982	0.002	0.943
Attendance on other related class	0.017	0.724	0.024	0.614	0.024	0.617	0.019	0.729
Total attendance	0.031	0.608	0.048	0.434	0.066	0.292	0.082	0.261
Change in attendance	0.103	**0.017**	0.13	**0.003**	0.124	**0.005**	0.13	**0.009**
Development 1 grade	-0.029	0.112	-0.04	**0.031**	-0.037	**0.051**	-0.036	**0.081**
Development 2 grade	0.006	0.737	0.004	0.825	0.001	0.94	0.003	0.882
Agriculture economics grade	-0.035	**0.006**	-0.044	**0.001**	-0.044	**0.001**	-0.041	**0.012**
Cumulative GPA	0.055	0.165	0.02	0.618	0.021	0.615	0.057	0.233
English result in ESLCE	-0.002	0.218	-0.003	0.122	-0.002	0.343	-0.003	0.247
English result in ten	-0.02	0.304	-0.011	0.58	-0.004	0.847	0.005	0.853
Math result in ESLCE	-0.002	**0.066**	-0.003	**0.052**	-0.002	**0.068**	-0.003	**0.043**
Math result in ten	0.066	**0**	0.056	**0**	0.057	**0**	0.055	**0.004**
Average ESLCE result	0.009	**0.01**	0.01	**0.004**	0.009	**0.012**	0.01	**0.018**
Share of A"s in ten	-0.049	0.572	-0.017	0.848	0.003	0.977	0.08	0.477
Share of B's in ten	0.037	0.558	0.045	0.48	0.066	0.322	0.137	**0.071**
Sex	0.024	0.529	0.025	0.513	0.017	0.656	0.009	0.827
Class level mark of the question	1.121	**0**	1.131	**0**	1.139	**0**	1.167	**0**
Afar	-0.176	**0.07**	-0.187	**0.056**	-0.168	**0.088**	-0.12	0.243
Amhara	0.033	0.219	-0.015	0.637	-0.01	0.737	-0.04	0.33
Oromia	-0.035	0.102	-0.033	0.117	-0.03	0.156	-0.025	0.318
Somalia	-0.017	0.7	-0.014	0.752	-0.008	0.853	0.025	0.626
Benishangul-Gumuz	0.139	**0.032**	0.13	**0.055**	0.135	**0.046**	0.15	**0.029**
SNNPR	-0.027	0.439	-0.029	0.425	-0.028	0.439	-0.007	0.858
Harari	-0.227	**0.011**	-0.225	**0.011**	-0.228	**0.01**	-0.185	**0.047**
Addis Ababa	-0.049	0.638	-0.047	0.655	-0.052	0.622	-0.028	0.8
Dire Dawa	-0.023	0.626	-0.012	0.796	-0.015	0.747	0.025	0.651

Dropping students which are having higher similarity with others and when they also having below 3 CGPA did not change much of the conclusion. When students with more than or equal to 80% of similarity and those with more than or equal to 70% similarity are dropped, only the positive impact of having good grade in development economics one become significant and there is no change in other effects. And if we restrict the similarity to below 60%, good performance at 10^{th} grade courses as measured by share of B's in all course is having positive impact. And the impact of regional location being in Afar becomes insignificant. But all other relationships remained more or less the same.

In general the results of the paper are more or less robust for any possible distortion that can be resulted for wide spread cheating. Means assuming that exam performance is directly related to knowledge will not be wrong since any possible cheating is not introducing any significant change in the conclusion. So the probability of cheating in the rural development objective exam is not expected to be zero, though very low; how ever what ever is the level of cheating, it was not wide spread to introduce inconsistency in the model estimation and conclusion.

4. Conclusion and recommendation

The paper was started with underlining hypothesis and assumption that 'high school does not matter'. Observing the wide spread poverty in the country and the more or less egalitarian quality of life enjoyed by university students the researcher was in firm believe that high school performance does not matter. But clear analysis of the data did show that high school performance and all related variables did matter especially for freshman students. But in line with wider empirical evidence the importance of high school performance did decline with time. So even though it can't be clear, if the level of understanding on high school courses or other correlated variables like socio economic environment of the family are exerting impact, it is clear that those who perform well in high school have higher probability that they will also perform well at university level. So the next research should try to see, if knowledge gained on high school is making university an easy place for good performance or the socio economic environment which resulted on good high school performance is also exerting effect at university level, too. But part of the 'myth' which stayed myth is the assumption that good performance on high school English and math are what are needed for good performance at university level. And any evidence can't be found to support such assumption. For the specific course even though high school performance was positively related to good performance its impact was simply marginal.

In general what is important is to notice that how important high school performance is, still it does not explain 100% of the variation on students' performance. Means high school performance and all related variables at best explain around 40-50% of the variability on university performance at freshman level. Moreover the importance will decline to 20 to 30 as student progress over time. This implies two things first high school performance is not the only factor for success means there is more than half of the variation on performance which can't be explained by high school performance and related variables. Means there is work that can be done at university level accepting the

high school level as given. And these variables which unrelated to high school are at least as important as high school performance. And this more the case given what is found is correlation not causation. The positive causation to ward one course's performance was marginal but the author is restrained from inferring from one course to the rest of the courses given the high level of correlation observed specially for freshman students.

Second it may be productive if some form rehabilitation and affirmative action is used to improve poorly performing students. This can include improving pedagogical knowledge of the lectures, following high school type of lecturing at least at freshman level and providing some basic high school level course at university for freshman students. The assumption behind such conclusion is that the reconditioning needed to make the students use their full capacity needs structural intervention to attain it in first or second semester of first year. The logical flow of the conclusion is that based on the assumption the students who are performing poorly are not necessarily low IQ students. They lack basic back ground but most importantly self confidence which is resulting on thier poor performance. This may not be reversed in just one or two semester unless structural and pragmatic intervention is done. And even though the researcher does not think the courses covered in university level are not complicated enough for any one with average IQ to understand them. Still following the theory of Zone of proximity development of Vygotsky (Elliot et al. 2000), there can be a lot of work that can be done to facilitate and stimulate the development of potential and the utilization of full potential even for those with capability.

As is documented by Wößmann (2003) in his international study which encompass 39 countries and more than 200 000 students it is the creation appropriate institution which create the necessary incentive for every one (student, teacher and others) to act optimally that matters most. So appropriate institutions must be putted in place to make sure students, lecturers and all supporting staff are working at their best. But most importantly the pedagogical tools which can improve effectiveness needed to be provided not by just another instructor with Degree in pedagogy but by those with good experience on teaching and pedagogical research. Inclination of the student is an important factor which means students perform well if they like the subject matter. So than pushing for strict attendance of students it may be much better to make the course more interesting and the exam more challenging. And the last but not least conclusions that have to be drawn from the study is that the quality of input provided to the students is an important factor on explaining performance as the poor quality materials that are provided are observed to have insignificant impact on performance.

But all the conclusions have to be accepted with reasonable reservation because the sample is not representative to all Ethiopian university students. So it has to be proven by different similar researches before any conclusion can be accepted. And the relatively more representative correlation analysis though established strong association it was not adequate enough to establish causation. Still formal and representative analysis is needed to show the level of causation behind mere association observed above. So the real value of this paper is to start the discussion than to draw the dividing line between fact and myth.

Notes

[1] See Mathewos 2000, Driscoll et al. 2008, Bacolod et al. 2006, Sander 1999, Beutel et al. 2002 and Wößmann 2003 for example

[2] Subjective question here means questions which evaluate the economic logic than the conclusion. Say you ask a student if free market is applicable to Ethiopia. And both those who say yes and no can be right or wrong based on their logic. Similarly indirect questions will ask for a question which needs critical thinking than mere reading. Good performances on both indirect and subjective questions are reflection (not determinant) of student knowledge. Both are observed in subjective exams and remember subjective exams and subjective questions are not the same (according to the definition used here).

[3] The interest of the student to ward the over all course is imperfectly controlled by using students' performance in related course.

[4] If any independent variable is not in one of the equations, its related coefficient will be zero in $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$.

[5] This assumption is redundant as the intercept will correct for non zero mean of the error term $\mathcal{E}$

[6] The first derivative of the standard deviation function or $h'(\boldsymbol{\alpha}\mathbf{Z})$ is a constant which can be ignored with out lose of any important information in the artificial regression.

[7] One section is composed of extension (evening class) students and there is no high school level data for them. The other two sections are regular students which attend class in normal class hour

[8] Which is help full in 2 to 3 chapters and is book focused in teaching rural development from India prospective

[9] Which is prepared for distance students some how mirroring the text book

[10] In other course lectured by the same instructor effort was made to collect data but most of them give the highest mark not because the lecture was perfect but they are afraid of any possible consequence. So unless some form of confidence is built the collection of such important data is impossible

[11] To give some examples Stephenson et al. (2006) based on Virginia Tech; Nolan et al. (2007) based on Sydney University and Salahd et al. (2005) based on University College of Medicine, Lagos Nigeria found that students which perform better in high school did perform better in university, too.

[12] And this is also well documented in series of studies by Nolan et al. (2007), Salahd et al. (2005), Fentaw (2001) and so on.

[13] And as we will see later on students performance is highly related to family income, education status of family head and gender of the student, among other things.

[14] Legally this is acceptable reason for missing class for with in limit. In practice it is the most given reason by all students in case they miss class

[15] Of course if the concept is repeated the lecture will be less detail

[16] See Mathewos 2000, Driscoll et al. 2008, Bacolod et al. 2006, Sander 1999 and Wößmann 2003, for example.

[17] And this is backed by study of Mathewos 2000, Bacolod et al. 2006, Beutel et al. 2002, Wößmann 2003 and others.

[18] Almost all of the papers in the reference have some contribution in some or all of the arguments and counter arguments.

[19] The right way to approach it is to fix all other variables at their average and to calculate the change in probability that can be resulted from change on the binary variable.

[20] actually the paper was initially titled as "myth and facts of university performance" with objective of showing unfounded-ness of the 'myth'

[21] The reason related factors are stated above is because correlation does not imply causation but simple association

[22] The finding is at high school level

Appendix
Algorithm for identification of maximum similarity in answer

In order to identify the maximum similarity the following algorithm is used. This will calculate student by student the maximum similarity between each student and all other students.

```
/*before starting all students' name or id will putted in one column
  and their answer to each question in numeric form will be putted in
  one column for each question  here there are 40 of them */
/*in this paper there are 120 students and 40 questions*/
/*some of the students and questions are dropped due to data problem in
  the main analysis. But there is no reason to drop them here*/
/*to use it for other data after arranging as above change the number
  after t and s below by related value to your data. The rest is the
  same*/

gen local t = 120 /*this is for total number of students*/
gen local s = 40 /*this is for number of questions*/
gen no = _n
gen tad = 0
gen similarity = 0
local i = 1
while `i' < (`t' - 1) {
local j = 1
while (`i' + `j') < `t' {
replace tad =  0  if  (no == `i' | no == (`i' + `j'))
local k = 1
while `k' < `s' {
if q`k'[_n + `i' - 1] == q`k'[_n + `j']{
replace tad = tad + 1
}
else {
replace tad = tad + 0
}
local k = `k' + 1
}
if similarity < (tad/`s') {
replace similarity = (tad/`s') if  (no == `i' | no == (`i' + `j'))
}
else {
replace similarity = similarity if  (no == `i' | no == (`i' + `j'))
}
local j = `j' + 1
}
local i = `i' + 1
}
drop tad

/* the values given in variable similarity on you data are the maximum
  similarity Between each students answer and any other student*/
```

Reference

1. Ahmed, Akhter U. and Mary Arednds Kuenning 2006, Do Crowded Classrooms Crowd Out Learning? Evidence from the Food for Education Program in Bangladesh, World Development, 34, 4: 665 – 684
2. Amare Sahile 2001, Effects of students' academic competence, self determination and motivation on school performance in Tana Haiq Sec. School, the Ethiopian journal of education, 21, 1
3. Anastasi, Anne 1961, Psychological testing, New York: The Macmillan Company,
4. Asadullah, Mohammad Niaz; Nazmul Chaudhurye and Amit Dar 2007, Student achievement conditioned upon school selection: Religious and secular secondary school quality in Bangladesh, Economics of Education Review, 26: 648–659
5. Asresash Demise, Ruth Shinebaum and Kassahun Melesse 2002, The problems of female students at Jimma University, Ethiopia: with some suggested solutions, Ethiopia Journal of Health Development, 16, 3: 257-266
6. Bacolod, Marigee P. and Justin L. Tobias 2006, Schools, school quality and achievement growth: Evidence from the Philippines, Economics of Education Review, 25 : 619 – 632
7. Beutel, Ann M. and William G. Axinn 2002, Gender, Social Change, and Educational Attainment, Economic Development and Cultural Change, 51: 109 –134
8. Driscoll, Donna; Dennis Halcoussis and Shirley Svorny (2008), Gains in standardized test scores: Evidence of diminishing returns to achievement, Economics of Education Review, 27: 211–220
9. Elliott, Stephen N., Thomas R. Kratochwill, Joan Littlefield cook and John F. Travers 2000, Educational Psychology: Effective teaching, Effective Learning, Boston: McGraw-Hill Higher Education
10. Fentaw Abegaz 2001, Admission type and female students' academic success at AAU: the case of social science students, The Ethiopian journal of education, 21, 1
11. Greene, William H. 2000, Econometric Analysis, 4th edition, New Jersey – USA: Prentice Hall, Inc.
12. United Nations Development Program (UNDP) 2007/8, Human Development Report, New York : Oxford University press
13. Herbert, Jennifer and Deborah Stipek 2005, The emergence of gender differences in children's perceptions of their academic competence, Applied Developmental Psychology, 26: 276 – 295
14. Mathewos Tamiru 2000, Factors affecting Academic achievement of students at bole senior secondary school, The Ethiopian journal of education, 20, 1
15. Nairne, S. James 1996, Psychology: The Adaptive Mind, Pacific Grove: Brooks/ Cole Publishing Company
16. Nolan, Elizabeth and Fredoun Z. Ahmadi-Esfahani 2007, Predicting performance in undergraduate agricultural economics, The Australian Journal of Agricultural and Resource Economics, 51: 1–15

17. Salahdeen H. M. and Murtala B. A. 2005, Relationship between Admission Grades and Performances of Students in the First Professional Examination in a New Medical School, African Journal of Biomedical Research, 8: 51 - 57
18. Sander, William 1999, Endogenous expenditures and student achievement, Economics Letters, 64 : 223–231
19. Santibañez, Lucrecia 2006, Why we should care if teachers get A's: Teacher test scores and student achievement in Mexico, Economics of Education Review, 25 : 510–520
20. Stephenson, Kurt; Anya McGuirk, Tricia Zeh, and Dixie Watts Reaves 2006, Comparisons of the Educational Value of Distance Delivered versus Traditional Classroom Instruction in Introductory Agricultural Economics, Review of Agricultural Economics, 27, 4 : 605–620
21. Wößmann, Ludger 2003, Schooling Resources, Educational Institutions and Student Performance: the International Evidence, Oxford Bulletin of Economics and Statistics, 65, 2
22. Verbeek, Marno 2006, A guide to modern econometrics, 2nd edition, Southern Gate, Chiehester – England: John Wiley & Sons, Ltd